This README file was generated on 2015-03-27 by Zewei Song

GENERAL INFORMATION

1. Title of Dataset: Effort Versus Reward: Preparing samples for fungal community characterization in high-throughput sequencing surveys of soils

2. File Information:

A. Filename: Representative OTU sequences.fa

Short description: representative sequences of OTUs in the OTU table.

B. Filename: Supplemental File S2. OTU table and heatmaps.xlsx

Short description: Raw OTU abundance data for different sites with associated data dictionary

C. Filename: Supplemental File S1. Diversity Indexes.xlsx

Short description: Breakdown of OTU diversity index grouped based on various experimental characteristics, such as extraction volume, extraction method, etc.

3. Principal Investigator Contact Information

A. Name: Linda Kinkel

B. Institution: University of Minnesota

C. Address: 1991 Upper Buford Circle, St Paul, MN, 55108

D. Email: kinkel@umn.edu

4. Associate or Co-investigator Contact Information

A. Name: Zewei Song

B. Institution: University of Minnesota

C. Address: 1991 Upper Buford Circle, St Paul, MN, 55108

D. Email: songx208@umn.edu

5. Date of data collection: 2013 - 2014

6. Geographic location of data collection (where was data collected?):

1) University of Minnesota Cedar Creek Ecosystem Science Reserve (CCR, 45°24'13" N, 93°11'20" W), Minnesota, 55301

2) University of Minnesota Cloquet Forestry Center (CFC, 46°40'45" N, 92°31'08" W), Minnesota, 55720

8. Information about funding sources that supported the collection of the data:

- USDA 2011-67019-30200

- University of Minnesota MnDrive

- National Science Foundation Grant EF 12-41895

- National Science Foundation Long-Term Ecological Research Network 0620652

METHODOLOGICAL INFORMATION

1. Description of methods used for collection/generation of data:

We investigated the effects of four procedural modifications to library preparation for high-throughput sequencing (HTS). The following treatments were considered: 1) the amount of soil used in DNA extraction, 2) the inclusion of additional steps (freeze/thaw cycles, sonication, or hot water bath incubation) in the extraction procedure, 3) the amount of DNA template used in polymerase chain reaction (PCR), and 4) the effect of sample pooling, either physically or computationally.

Soils were collected from two study sites in Minnesota. Four replicate soils were collected in each site, and were also combined to form a physical bulked soil sample. These samples were applied to the treatments described above. The DNA library was sequenced using an Illumina MiSeq paired-end 250 base pair. An operational taxonomic unit (OUT) table (filename: OTU table and heatmaps.xlsx) was generated after quality trim and OTU clustering. All analyses in our submitted manuscript were based on OUT table and heatmaps.xlsx and diversity indexes (filename: Diversity Indexes.xlsx) calculated from it.

2. Quality-assurance procedures performed on the data:

The raw sequence data was trimmed to remove any attached primer or adapters. The data was then quality filtered to get rid of low quality bases, most them at the 3’ end. The data after quality control had a minimal quality score at 20 (1% error rate at a single base).

3. People involved with sample collection, processing, analysis and/or submission:

Zewei Song[1*], Dan Schlatter[1], Peter Kennedy[2], Linda L. Kinkel[1], H. Corby Kistler[1,3], Nhu Nguyen[2], and Scott T. Bates[1]

[1]Department of Plant Pathology, University of Minnesota, Saint Paul, MN 55108, USA

[2]Department of Plant Biology, University of Minnesota, Saint Paul, MN 55108, USA

[3]USDA ARS Cereal Disease Laboratory, Saint Paul, MN 55108, USA

DATA-SPECIFIC INFORMATION

Supplemental File S2: OTU table and heatmaps.xlsx:

Sample format of sample label in OTU tables:

CCR_S1_0.25g_Sonication_20ng

1. First part (CCR) refers to site of sample collection
    - CCR: Cedar Creek
    - CFC: Cloquet Forest
2. Second part (S1) refers to sample type
    - (1-4): soil samples from the individual plots
    - Physical: bulked soil from the four individual plots
    - Computational: in silico combination of the four plot sample
3. Third part (0.25g) refers to amount of soil used in DNA extraction
4. Fourth part (Sonication) refers to modification used during extraction of soil DNA
5. Fifth part (20ng) refers to amount of template used in PCR

Supplemental File S1: Diversity Indexes.xlsx:

Description of the four sheets:

1. Extraction Volume: Amount of soil used in DNA extraction (10g, 1g, or 0.25g as control)

2. Extraction Method: Modification to the original DNA extraction protocol (freeze/thaw, sonification, water bath, and control)

3. Template Amount: Amount of DNA template used in PCR (10ng, 20ng as control, and 40ng)

4. Soil Bulking: Bulked soils (combination of the four individual plot samples) compare to individual soil (Sample 1-4, or Bulk)

Column values:

1. Soil Sample: Site (Cedar Creek for CCR, and Cloquet Forest for CFC) and plot number (1-4)

2. Treatment: library preparation procedures

3. Richness: Number of OTUs in the sample

4. Simpson: alpha diversity index (see: http://en.wikipedia.org/wiki/Diversity_index)

5. Shannon: alpha diversity index (see: http://en.wikipedia.org/wiki/Diversity_index)

6. Chao1: Estimate of species richness (see http://www.mothur.org/wiki/Chao)

7. Soil Amount: amount of soil used in DNA extraction

8. Kit Type: kit used to extract DNA from sample (PowerSoil Kit or PowerMax Kit)

Description of Treatments:

freeze/thaw: The freeze/thaw treatment consisted of three rounds of freezing and thawing by submersing tubes in liquid N_2 for 5-10 seconds followed by an incubation step in a 65°C water bath for 10 min.

sonification: The sonication treatment consisted of sonicating tubes for 10 min in a Branson 8200 sonicator (Thomas Scientific, Swedesboro, NJ, USA).

water bath: The heating treatment consisted of incubating tubes in a 65°C water bath for 10 min.

control: Standard protocol of PowerSoil or PowerMax Soil.

Description of Kit Types:

PowerSoil Kit: A standard soil DNA extraction kit form MO BIO (http://www.mobio.com/soil-dna-isolation/powersoil-dna-isolation-kit.html). It uses 0.25 gram of soil.

PowerMax Soil Kit: A similar kits as PowerSoil, but uses 10 gram soil instead (http://www.mobio.com/soil-dna-isolation/powermax-soil-dna-isolation-kit.html)

SHARING/ACCESS INFORMATION

1. Licenses/restrictions placed on the data:

    - CC0 1.0 Universal – Public Domain Dedication

2. Links to publications that cite or use the data:

    - One manuscript in revision